

Was Kant a Normativist or Naturalist for Mental Illness?

Dominic A. Sisti PhD

University of Pennsylvania

Department of Medical Ethics & Health Policy

Philadelphia, PA, USA

ABSTRACT

Kant's philosophy of medicine and biology provides insight into contemporary debates about the concepts of health, disease, and illness. In particular, Kant's writings on mental illness offer a unique form of naturalist nosology and help add further relief to the question about whether certain mental illnesses should be considered natural kinds, social artifacts or a hybrid of both. In this paper, I survey some of Kant's writings on mental illness and then link his account to contemporary perspectives. I argue that Kant offers a unique form of nonnormative nosology that is distinctively nonreductionist.

Key Words: Kant, mental illness, health, disease, anthropology, nosology

Introduction

Kant's work in the philosophy of medicine and biology, particularly his notions of mental health and illness, provides a precursor to the ongoing debate about what should count as a disease within psychiatric nosology. In this paper, I wish to draw primarily on Kant's *Anthropology* and his *Critique of the Power of Judgment* to show that mental illnesses might result from a derangement of various cognitive faculties such as reason, understanding, and judgment. I therefore argue that for Kant, mental illness can be seen as a fundamental problem in a person's ability to reach and exercise their ultimate end—living an autonomous and morally good life.

Contemporary theories on the concepts of mental health and illness can be categorized into three broad groups: naturalism, normativism, and hybrid models (Ereshefsky, 2007). Naturalist theories describe mental disorder in biological terms and ground normal functioning as that which is species typical. Diseases, it is argued, can be ascertained independent of value judgments—they are truly biological dysfunctions. In contrast, normativists view diseases as being imbued with social and political values. In addition to developing sophisticated epistemic claims about medical and biological knowledge, normativists hold up cases of value-laden syndromes of the past to make their case that it is

impossible to label particular behaviors as dysfunction without reference to some value-based reference point. Hybrid models attempt to provide the best of both worlds—they accept biological brute facts as such, but maintain that social values largely determine which biological abnormalities are deemed harmful.

I argue that Kant's concept of mental illness is constructed on what might now be considered naturalist criteria. However, his concept of mental illness cannot be assessed by purely scientific means because, as we will see, it relies on a teleological judgment of the ends of persons. I therefore wish to propose that Kant offers a novel kind of naturalist nosology for mental disorder that is distinctively nonreductionist and which has not been sufficiently examined in the recent philosophy of medicine and bioethics literature. I will contrast Kant's concept of mental illness with that of Jerome Wakefield, whose harmful dysfunction hybrid model of illness, although different in important respects, resonates with Kant. Wakefield's model will also serve as a template against which we can begin to examine Kant's theory in the context of contemporary philosophy of psychiatry. Let us first survey the way in which Kant describes mental illness empirically and then look at how mental illness is understood within Kant's rational psychology.

Kant on Mental Illness: Empirical Thumbnails

During the critical period, Kant clearly rejected the philosophical promise of rational psychology as a way to fully understand the human mind and all its faculties. He sets out this refutation explicitly in the *Critique of Pure Reason* in the section on the paralogism of pure reason, which covers rational psychology. Here Kant (2007) argues that rational psychology is separate from his transcendental project, as it can never truly address the central transcendental proposition, "I think". Because rational psychology is ill equipped to explain the *cogito*, empirical psychology—which is seemingly based on rational psychology—should be inadequate as well.

It is therefore somewhat surprising that Kant's treatment of mental illness is almost exclusively empirical and a matter of 'pragmatic anthropology'. We should note, as Frieson (2009) points out, Kant's empiricism is not of the experimental variety, but rather is observational and anecdotal. His pragmatic anthropology stands

in stark contrast to the more popular rationalist undertakings of Kant and his contemporaries, setting it apart for being immediately applicable to and reflective of the real world.

Although empirical discussions of mental health and illness do not figure prominently in Kant's main philosophical corpus, we do find discussions of mental illness and, particularly, hypochondria in several of Kant's pre-critical and lesser-known writings (Shell, 1996). Kant, himself disposed to hypochondriasis, explicates the various kinds of mental illnesses starting from his own personal experiences. For example, in *Investigation Concerning Diseases of the Head* (Kant, 1764) as a preview of the disease of *melancholia*, Kant discusses the distinction between the delusions caused by hypochondria and those of other mental illness. This distinction becomes formalized, as we will see, in *Anthropology*.

In another instance, in his satirical *Dreams of a Spirit-Seer* (1766), Kant rebukes the enthusiast Swedenborg by essentially labeling his mystical experiences the ravings of a lunatic, who confuses the particular 'nerve vibrations' of fantasies—which reside internally—with real, externally caused sensations:

The peculiarity of this disease is that the confused individual places mere objects of his imagination outside of himself, and considers them to be real and present objects. (pg. 80)

We also find a discussion of mental illness in Kant's *Rektoratsrede*, which was probably delivered in 1788 and entitled—"On the Philosopher's Medicine of the Body." In this lecture, Kant first articulates the "philosopher's regime," which was Kant's way of treating ill predispositions, such as his own hypochondria, and mental illnesses through dietetics, the redirection of one's attention toward less worrisome thoughts and through play and intentional distraction (Shell, 1996).

Though in Kant's model it seems insanity can be prevented through the philosopher's regimen, it also seems that once a person has gone mad, they need to be treated medically. In fact, as Shell (1996) notes, Kant recognizes a hereditary "seed" for insanity that germinates during adolescence and is largely irrepressible once it "breaks out." Kant is pessimistic about the efficacy of medical treatments for full-blown mental illness. Additionally here, we see hypochondria take its central place in Kant's nosology. Later, *hypochondria generalis* (or unlocalized hypochondria, characterized by worry and anxiety) appears as a species of mental illness within the larger category *melancholia*.¹

In another episode, as Shell (1996) recounts, Kant's *Nachlaß* contains descriptions of Kant's notion of mental illness as imaginary freedom. Here he describes hypochondria as a disease of the imagination, arguing that mental illness can be rooted in the mistaken feeling that one is free when they actually are not. Kant illustrates this kind of delusion of freedom by way of metaphor, describing the relationship between women and men where a woman "governs man by letting him imagine himself her master and the prince who gives his subjects" imaginary freedom to enhance power (Shell, 1996, pg 292). In theoretical contrast, below we will see how a misunderstanding of one's true freedom becomes inevitably a frustration in the actualization of one's ultimate end and is the basis for a psychological disorder.

One final empirical anecdote is worth mentioning. We find in the

third and final section of *Conflict of the Faculties* (1798) Kant's essay, "On the power of the mind to master its morbid feelings by sheer resolution." Here Kant further develops, through personal disclosure, his understanding of his own hypochondria, which he describes as the exact opposite of the mind's power to master its pathological feelings. In a way, Kant's practice of the philosopher's regimen is a sort of auto-experiment in the prevention of both mental and physical illness, for which he provides this report. Kant talks in some detail about his own physical symptoms such as catarrh and insomnia and he outlines a particular breathing exercise as a remedy. He then goes on to describe how pathological feelings can take over at inopportune times as a result of unhealthy habits. For example, he describes how "reflection," which I read as rumination, during mealtime results in hypochondria or vertigo.

It is important to note that Kant's hypochondriacally motivated habits do not altogether count as a biologically based mental illness. Kant indicates that neither hypochondria nor melancholia have physiological causes, but there are physiological factors that make one susceptible to both. These susceptibilities can be overcome, and we will see that melancholia and hypochondria are, if controlled, favorable personal attributes.

Mental Illness in *Anthropology*

The above anecdotes provide a glimpse into Kant's overall understanding of mental health and illness. A systematic treatment can be found in Kant's *Anthropology from a Pragmatic Point of View* (1798). In *Anthropology* we find the empirical bridge between inner experiences of human reason and understanding and outward behaviors, which are indicative of either psychological health or illness. In the first part of the book, Kant describes how the cognitive faculty provides various other capabilities and differentiates between the lower and higher cognitive faculties. The former is the recipient of sense data and autonomic response; the latter is the locus of reason and thoughtful choice. In Kant's words, the higher faculty is characterized by,

spontaneity of apperception, that is, of pure consciousness of the activity that constitutes thinking. It belongs to logic (a system of rules of the understanding), as the [lower faculty] belongs to psychology (a system of all inner perceptions under laws of nature) and establishes inner experience. (pgs. 29-30)

Kant then describes the functions of each faculty. The lower faculty senses passively, while Understanding orders the "inner appearances" perceived by the lower faculty. Interestingly, Kant claims that the sense faculty is misunderstood as something that plays an active role in muddling the understanding. On the contrary, the senses are passive, having no power to confuse, deceive or order the understanding. Sensibility, Kant says, is present in irrational animals, which can manage it provisionally, "following implanted instincts like a people without a sovereign." (pg. 90) In fact, the key cognitive role of the senses is to remain at the "disposal of the Understanding." Confusion about the roles of each faculty results when one thinks that an intellectual insight springs directly from the inner sense (without the help of understanding) (pg. 36). Kant describes this as an instance of a derangement of the senses similar to "enthusiasm," reminiscent of *Dreams* decades earlier in which Kant labels Swedenborg's enthusiasm as lunacy.

The overarching power of Understanding (i.e. the entire higher faculty) consists of more specific elements of understanding, the power of judgment (*iudicium*), and reason. Kant then claims that the healthy mind is composed of *correct* understanding, *practiced* judgment, and *thorough* reason. For our purposes, we might conclude that the inability to differentiate between primitive sensual impressions and well-formed thoughts derived from human understanding may serve as a very rough criterion for Kant's notion of mental illness. The extent to which this confusion results in the degradation of freedom should then serve as a benchmark for the severity of the illness.

For Kant, the proper use of the understanding is, by definition, "healthy understanding." However, the faculty of understanding can be short-circuited through a lack of "precision" and "sufficiency," which results in a misunderstanding of the "appropriateness" of a concept. Kant thus defines ideal understanding in terms of a parsimonious explanation for whatever the situation is at hand. It turns out that understanding can be improved and sharpened through education. In contrast, the power of judgment can only be improved through experience and maturity. "Reason" in this context is essentially the same as in Kant's moral philosophy. "Reason is the faculty of deriving the particular from the universal and thus of representing it according to principles and as necessary" (pg. 93). This invocation of moral reason foreshadows what we might consider to be the unique element of Kant's nosology: that the concept of mental illness is uniquely nonnormative because it is in turn a measure of the extent to which human reason is employed in accord with universal, objective principles.

In *Anthropology*, Kant's nosology is composed of two forms of mental illnesses: *melancholia* and mental derangement (pg. 106). *Melancholia* is the anxiety that results from "certain internal physical sensations," which in the case of hypochondria, compel one to worry about their health and their life "business" more generally. "Anxious fear, childish in character, of the thought of death nourishes" hypochondria (pg. 107). It is thus the inclination to worry, to "self-torment," and to suffer from mood swings that characterizes the melancholic-hypochondriac. Because melancholia is a milder form of mental illness and seems to be remediable by the philosopher's regime, we might say this set of illnesses corresponds to what we now call neuroses.

In contrast, derangement, the second major division of mental illness, is much less amenable to correction or cure. Here Kant seems resigned to the fact that derangement is untreatable and he even hesitates in his analysis of it because he takes such study to be of little use:

It is difficult to bring a systematic division into what is essential and incurable disorder. It is also of little use to occupy oneself with it, because all methods of cure in this respect must turn out to be fruitless, since the powers of the subject do not cooperate (as is the case with bodily diseases), and yet the goal can only be attained through his own use of understanding (pgs. 108-109).

Thus, we might liken Kantian derangements to severe psychotic disorders, certain personality disorders or other egosyntonic mental illnesses where the patient lacks insight into her illness.

Within the rubric of derangement, Kant describes three

broad categories of dysfunction (tumultuous, methodical, and systematic), which are cross-categorized into the following particular disturbances. The first kind of derangement is *amentia* (*unsinnigkeit*) or the inability to bring coherence to one's thought and experience. This is a derangement of the tumultuous kind. In a sexist slight, Kant says that "in lunatic asylums it is women who, owing to their talkativeness, are most subject to this disease: that is, their lively power of imagination inserts so much into what they are relating that no one grasps what they actually wanted to say" (pg. 109).

Second, Kant describes *dementia* as a kind of incurable methodical derangement in which a person mistakes self-made representations as perceptions. Such may be the case in those who are paranoid and delusional. The demented person, "believe they are surrounded by enemies everywhere, [they] consider all glances, words, and otherwise indifferent actions of others as aimed against them personally and as traps set for them." (Kant, 2006 p. 109). Fortunately the demented are not a threat to others, leading Kant to say they "do not need to be locked up for reasons of safety."

Third, *insania*, Kant says "is a deranged *power of judgment* in which the mind is held in suspense by means of analogies that are confused with concepts of similar things, and thus the power of imagination, in a play resembling understanding, conjures up the connection of disparate things as universal, under which the representations of the universal are contained" (Kant, 2006, p. 109). This is an interesting form of mental illness, because it directly relates to the power of judgment, which we will more closely examine below.

The fourth is *vesania*, which Kant defines as the "sickness of deranged reason" (Kant, 2006 p. 110). This is the most fundamental and extreme form of mental disorder, obfuscating basic *a priori* truths wherein the patient thinks they can square a circle, invent a perpetual motion machine, or comprehend the mystery of the Trinity. Kant says that *vesania* is caused by positive unreason—where up is down, black is white and seven plus five equals eleven. All forms of Kantian madness share one common characteristic: the patient lacks common sense and instead relies on a sense of private logic. It is therefore clear, from Kant's empirical account in *Anthropology*, that mental illness is the product of a short-circuiting of the cognitive faculties, with judgment ceasing to serve as the reliable linchpin of the regulative system that includes understanding and reason.

Toward the end of *Anthropology*, Kant provides an analysis of how we might understand the interior of the person from their exterior appearance. This analysis further reveals Kant's overall understanding of mental health as moral health. We find this in Kant's description of *character* and its constitutive *temperaments*, each of which reveals the unique way in which a person thinks. The four classic temperaments are divided into two groups: the temperaments of feeling and the temperaments of activity. The two temperaments of feeling are the sanguine and the melancholy; the two of activity are the choleric and phlegmatic. Kant goes into the relationship among the four temperaments, such as how they complement and how they cancel one another out.

For our purposes, and as mentioned above, it is important to note that the temperaments determine a person's predisposition for

a particular illness, but are otherwise harmless and might even be beneficial. For example, though a person with melancholy temperament is disposed to the illness of melancholia, she actually thinks deeply in contrast to the shallow nature of the sanguine person. All of these features make up a person's empirical character, which in turn ultimately determine what a person "makes of himself" (Kant, 2006 pg. 192). Thus, a key facet of Kant's anthropology of mental illness is that it is teleological. As Frierson (2009) notes, this teleological feature is apparent in the way Kant describes the ideal character and purposeful aim of humankind to cultivate, civilize, and moralize itself. Kant's third *Critique* provides the key to understanding how teleological judgment supports his concepts of mental health and disorder.

Teleological Judgment, Reason, & Mental Health

In the first part of the *Critique of the Power of Judgment*, Kant describes how aesthetic judgment is part and parcel of an overall assessment of an object's form and function—that is, of an object's teleological nature. We are more interested in the second part of the *Critique of the Power of Judgment*, where Kant details further his concept of teleology vis-à-vis living organisms. Kant's concepts of teleology and teleological judgment are related to his concepts of mental health and illness for two reasons.

First, teleological judgment is important because it is a special form of judgment that transcends the limits of mechanistic or scientific inquiry. Second, once it is fully demonstrated that all life is purposive, we find that proper functioning, including cognitive functioning, leads toward an organism's teleological end. Accordingly, any frustration in the fulfillment of that end should be considered a dysfunction. In the case of persons, the upshot is that the ultimate end is autonomy and living the moral life.

The first dimension of teleological judgment is its reflective power in judging final causes. The reflective power is necessary for negotiating the dialectic expressed in the antinomy of the power of judgment:

Thesis: All generation of material things is possible in accordance with merely mechanical laws.

Antithesis: Some generation of such things is not possible in accordance with merely mechanical laws (Kant 2000, p. 259).

The reflective power of judgment—or the ability to ascertain the connection of all things in accordance with final causes—is a transcendental power and it supersedes the determinative power of judgment, which is limited to the capacity to recognize empirical realities (20:211). To engage in reflective teleological judgment is to observe the natural world, in all its beauty, schematizing the world according to a system of interdependent parts that work according to the observable laws of nature. Kant most clearly states this point in his First Introduction to the *Critique*,

The special principle of the power of judgment is thus: Nature specifies its general laws into empirical ones, in accordance with the form of a logical system, in behalf of the power of judgment (pg. 19).

The second dimension of the power of judgment is simply that the design of life itself is fundamentally evolving. Kant here invokes an intelligent designer to explain the way in which evolution works: that organisms are systematically working toward a final cause set out by an "intelligent world-cause" [Kant, 2000, p. 260]. This is a heuristic that allows the investigator of nature not to work entirely in vain by allowing him to base judgments on a fundamental principle of original organization (Kant, 2000, p. 287).

Human beings are the only organisms with the ability to recognize both the order and *telos* of the natural world, and transcend that order through exercising freedom that comes from understanding. Kant says this explicitly in a long excerpt that, nonetheless, deserves a full presentation,

Now we have in the world only a single sort of beings whose causality is teleological, i.e., aimed at ends and yet at the same time so constituted that the law in accordance with which they have to determine ends is represented by themselves as unconditioned and independent of natural conditions but yet as necessary in itself. The being of this sort is the human being, though considered as noumenon: the only natural being in which we can nevertheless cognize, on the basis of its own constitution, a supersensible faculty (freedom) and even the law of the causality together with the object that it can set for itself as the highest end (the highest good in the world)... Now of the human being... as a moral being, it cannot be further asked why it exists. His existence contains the highest end itself, to which, as far as he is capable he can subject the whole of nature, or against which at least he need not hold himself to be subjected by any influence from nature [9, p. 302].

Thus human beings *qua* moral beings are, in a sense, the keystone species. All of nature is "teleologically subordinated" to the final ends of freedom and morality of humans. To summarize these two dimensions: the *telos* of an organism is its ultimate end for which it has been (hypothetically) designed to attain; to assess that end requires a special form of judgment that is equipped to recognize and understand the order and organization of complex systems and their ends.

Though there continues to be debate about the meaning of Kant's teleological argument, purposiveness in the context of natural design serves as a heuristic device that allows humans to understand the natural world. For Kant, teleology is not necessarily a representation of what is *truly* going on in nature; even if there were a design and designer of the world, it would be impossible for the human mind to fully understand it. And so, teleological judgment is a byproduct of human understanding and its need to look for purposiveness in the world—as Kant's treatment of aesthetics makes clear in the first part of the *Critique*. Likewise, in science, it is necessary to envision a systematic organization among parts of nature. From individual body parts working together upwards through individuals of a species, different species in a community, and communities within an ecosystem, all of life's levels of hierarchy and organization are artifacts of the biologists' (read: all persons') need and penchant for envisioning patterns set down by an overarching design. Thus, the ability to intuit an "intelligible substratum of nature" and recognize "intentionality is a product of our human minds" (Kant, 2000 pg. 275).

Because the ultimate end of human persons is freedom and autonomous living, mental illness—as defined in *Anthropology*—is a frustration of the attainment of this end. Moreover, as Frierson (2009) notes, specific natural activities (sleep, laughter) and mental illnesses within Kant's nosology are teleologically grounded. It seems reasonable to assert that any diminished capacity for teleological judgment that stands in the way of freedom, clouds understanding, or corrupts other forms of judgment maps onto Kant's definition of mental illness. For example, as a paragon of derangement, *insania* might be considered the hypertrophy of the power of teleological judgment—in which patterns, designs and causes are recognized in everything the insane person encounters but where there are in reality none.

Overall, it seems that mental illness in Kantian terms is two pronged: first, it upsets one's ability to appropriately apply the faculty of judgment, particularly when needed for reflective teleological judgment and, second, as a result of this frustration, mental illness prevents one from reaching his final end as an autonomous person. This dual structure will serve as a way to compare Kant's concepts with contemporary arguments about the nature of mental illness.

The Contemporary Debate: Situating Kantian Mental Illness

The contemporary debate about the nature of health and disease found in the philosophy of medicine generally pits two perspectives against one another. One group argues that diseases are biologically based and that health can be determined in terms of species typical functioning. The most well known contemporary philosopher of this group is probably Christopher Boorse (1975), who famously argued that diseases are value free concepts that are based on,

the autonomous framework of medical theory, a body of doctrine that describes the functioning of a healthy body, classifies various deviations from such functioning as diseases...This theoretical corpus looks in every way continuous with theory in biology and other natural sciences, and I believe it to be value-free (p. 55-56).

Boorse (1976) also argues that this function-based definition applies to diseases of the mind, and we can use his naturalism as the standard theory held by scientists and philosophers who believe that mental illness is reducible to dysfunctions of the brain. At its base, the naturalist position is teleological and is based on evolutionary biology. Cells, organs, systems, and individuals organisms function in accordance with a particular set of goals, namely survival and reproduction.

The second group argues that health and disease are normative concepts and that neither health nor disease can be fully defined in merely biological or evolutionary terms. Rather, health and disease are value-laden concepts and to understand them we must recognize and unpack the values that are smuggled into biomedical definitions of health and disease. Evidence from the history of medicine seems to support the normative position, in that we find many examples in which the concepts of health and disease are social-political constructs. Such was the case with *drapetomania*, or the disease of afflicted slaves who had a propensity to run away, and

the inclusion of homosexuality within the psychiatric nosology.

But while these historical episodes show that health and disease can be socially constructed, it does not follow that this is a correct way to characterize the ontological nature of *true* diseases. In fact, one might argue, recognition of the cases of pseudo-illness actually relies on an objective template against which such a claim can be made. In any case, the core debate about the concepts of health and disease pits the naturalists, who see disease in terms of deviations in evolutionary form and function, against the normativists, who see the concept of disease as always implicitly value laden. There are moderate or hybrid positions that try to bridge this philosophical gap—I shall mention Wakefield's below. But for now the above dichotomy will be helpful as we situate Kant's concept of mental illness within these broad categories.

It seems clear that Kant's concepts of mental health and illness are those of someone who today would be considered an unabashed naturalist. We do not have to dig too deeply to find evidence of this claim. In *Anthropology* we saw how mental illnesses are frustrations and disorders of natural faculties. Likewise, as in the third *Critique*, the faculties of the mind and brain (or in Kant's words, the head) are meant to serve the teleological ends of the individual person to live freely and rationally. Later in that *Critique*, Kant provides a systematic proto-evolutionary account of the goals of life (without natural selection) in which every organic system works in an organized fashion toward reaching life's final end, and this includes the brain and central nervous system. Therefore, armed with the empirical evidence Kant provides in tandem with his concepts of teleological judgment and evolutionary teleology, we can feel confident in claiming Kant is a naturalist for the concept of mental illness.

But now if we look a bit more carefully, Kant's version of naturalism is different from standard contemporary models such as Boorse's. For Kant, the concept of mental illness we learn about in *Anthropology* and his pre-critical writings seem to be characterized less by discrete biological dysfunctions. Rather, mental illness emerges from a pervasive lack of understanding about the proper goals of life. This makes Kant's concept difficult to precisely situate within the contemporary debate about the concepts of health and disease because, unlike Boorse's concept, it is distinctly nonreductionist and highly moralized. In fact, to try to reduce mental illness to a biological concept would require the use of scientific or mechanical reasoning, which Kant says is ill equipped for detecting a detour on the teleology of life. The moral dimension of Kant's concept is key to this unique kind of naturalism. With the failure of the faculties of reason, one fails to live according to the categorical imperative, and thus the mentally ill are doomed to a life of heteronomy.²

Could Kant Be a Moderate Normativist for Mental Illness?

There are grounds for an objection to the above analysis. Perhaps Kant's concept of mental illness is actually less objective than it seems. Is there, perhaps, room in Kant's version of mental illness for subjective values? It seems unlikely that one could mount a compelling defense of the claim that Kant is a pure normativist.

Such a claim would ultimately ignore the way in which Kant's concept of mental illness, as with all of Kant's concepts, is dependent upon the universal moral imperative. A more reasonable possibility might be that Kant is a moderate normativist for mental illness, and that he allows for some values to influence his concepts of mental health and illness. At base diseases of the mind are defined in terms of natural functioning, the ways in which mental dysfunction is harmful (or not) could be for Kant a value-based appraisal and an empirical question.

Such a contemporary hybrid model of disorder has been proposed by Jerome Wakefield (1992). For Wakefield, a disorder is defined in terms of two essential criteria. The first is that there is a dysfunction in the way an organ or organ system works. Second, the dysfunction is determined to be harmful. It is this second criterion that represents the overlaying of values to what is otherwise a purely objective phenomenon of dysfunction (i.e. the first criterion). Known as the harmful dysfunction concept of disorder, Wakefield has attempted to straddle the two sets of philosophical positions by offering a model that both accounts for the biological bases of mental illness (say, lesions on the brain) and social values that determine such biological realities to be bad. For Wakefield, a mental disorder is, on the basic level of biological functioning, an inability to fulfill the natural goals of the brain—for whatever reason—and is therefore considered to be a harmful disability. This seems to resonate with Kant's categorization of mental illness in *Anthropology*. Recall, above, each mental illness was described as general dysfunction (methodical, systematic, etc.) and was further specified in terms of the empirical harms or instances of personal devaluation caused by the illness.

What we see in Wakefield's harmful dysfunction model is his recognition that the ends of human life and the function of the mind are to be evolutionarily assessed. This is, for Wakefield, clearly possible through biological sciences. That fact that a scientific explanation of every mental disorder has not yet been found does not mean that it will not. In fact, Wakefield thinks that mental illnesses that lack a solid scientific footing, such as personality disorders, are initial estimations of what will be revealed through further study. But here Wakefield departs from Kant, who argues that insights about dysfunctions of the human mind ultimately rely on teleological judgment and will never be fully gleaned from scientific investigation. This is the key to why Kant's naturalism is different from all others: it is nonreductionist.

Interestingly, we find a rather similar model offered by Megone who defines mental illness as a deviation in the universal human function of reason. Megone, however, bases his model on Aristotelian metaphysics. He says that,

illness, whether bodily or mental, should be analyzed as an incapacitating failure of bodily or mental capacities, respectively, to realize their functions. It was suggested that the notion of function here should be understood as Aristotle had understood it (incorporating, in modern terms, an evaluative component): Human beings have a function in the sense that there are goals or purposes that good human beings will realize (actualize). The Aristotelian claim is that the ultimate goal for a good human being as a whole is *to live the life of a fully rational animal* (Megone, 2000 [emphasis added]).

Without going into too much detail here about Megone's model, it parallels Kant's overall understanding of the ends of humanity and how a disruption therein might be considered an illness. Thus, I believe Megone's model offers a promising hook for continuing a deeper Kantian analysis of the concepts of mental health and illness, which has been overlooked in contemporary philosophy of medicine. And without too much work, I think Megone's model could be adapted and enhanced by including elements of Kantian philosophy.

In summary, I argue Kant does not offer a hybrid account of mental illness, but rather a unique and fully naturalist account. For Kant, there really is no role for subjective values in determining mental illness. We come to recognize mental illness through our teleological judgment of the purpose of humankind, which is itself part of a transcendental psychology impervious to scientific explanation. Kant's concepts of mental illness and health are therefore fundamentally nonreductionist and nonnormative. This is evidenced in both his empirical examples of mental illness and part two of his third *Critique*.

Footnotes

1. For a complete taxonomy of Kant's concept of mental disorder, see Frierson (2009a).
2. Though Kant (2006) is describing sickness of the soul and not mental illness per se, the following reveals the importance of the healthy rational mind for full personhood: "Complete mental deficiency, which either does not suffice even for animal use or the vital force, or which is just sufficient for a mechanical imitation of external actions that are possible through animals (sawing, digging, and so on) is called idiocy. It cannot really be called a sickness of soul; it is rather absence of soul" (p.106).

References

- Boorse, C. (1975). On the Distinction between Disease and Illness, *Philosophy & Public Affairs* 5(1): 49-68.
- Boorse, C. (1976). What a theory of mental health should be, *Journal for the Theory of Social Behaviour*, 6(1): 61-84.
- Ereshefsky, M. (2009). Defining 'health' and 'disease'. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 40: 221-227.
- Frierson, P. (2009). Kant on mental disorder. Part 1: An Overview. *History of Psychiatry*. 20:267-289.
- Kant, I. [1764] 2011. *Essay on the Maladies of the Head*, in *Observations on the Feeling of the Beautiful and Sublime and Other Writings* eds. Frierson and Guyer, Cambridge: Cambridge University Press.
- Kant, I. [1766] 1900. *Dreams of a spirit-seer illustrated by dreams of metaphysics*, trans. E.F. Goerwitz. New York: The MacMillan Co. (Full text available at the following URL: <http://archive.org/details/dreamsofspiritse00kant>)
- Kant, I. [1798] 1996. On the power of the mind to master its morbid feelings by sheer resolution. In *The conflict of the faculties*. In *Religion and rational theology*, eds. Guyer and Wood, trans. Mary Gregor, 1979. Cambridge: Cambridge University Press.

- Kant, I. [1790] 2000. *Critique of the power of judgment*, trans. P. Guyer and E. Matthews. Cambridge: Cambridge University Press.
- Kant, I. [1798] 2006. *Anthropology from a pragmatic point of view*, ed. R. Louden. Cambridge: Cambridge University Press.
- Kant, I. [1781/1788] 2007. *Critique of pure reason*, eds., Guyer and Wood. New York: Cambridge University Press.
- Kuehn, Manfred. (2006). Introduction. In *Anthropology from a pragmatic point of view*. Ed., trans. Robert B. Louden. New York: Cambridge University Press.
- Megone, C. (2000). Mental illness, human function, and values. *Philosophy, Psychiatry, & Psychology*, 7(1):45-65.
- Shell, S.M. (1996). *The embodiment of reason: Kant on spirit, generation, and community*. Chicago: University of Chicago Press.
- Sturm, T. (2001). Kant on empirical psychology: How not to investigate the human mind. In *Kant and the sciences*, ed. Eric Watkins. New York: Oxford University Press.
- Wakefield, J.C. (1992). The concept of mental disorder. On the boundary between biological facts and social values. *American Psychologist*, 47(3):373-388.

Acknowledgments

I thank Fred Rauscher and Susan Meld Shell for their feedback on earlier drafts of this paper. Thanks also to John Maher, Jason Schwartz and reviewers of the Journal of Ethics in Mental Health. Katherine Buckley provided exceptional editorial assistance. This project was supported in part by the Scattergood Program for Applied Ethics of Behavioral Healthcare at the University of Pennsylvania Department of Medical Ethics & Health Policy.

Competing Interests: none

Address for Correspondence:

Dominic A. Sisti, Ph.D.
 University of Pennsylvania
 Department of Medical Ethics & Health Policy
 3401 Market Street, Suite 320
 Philadelphia, PA 19104-3308
 Tel. (215) 898-7136
 Fax (215) 573-3036

e-mail: sistid@mail.med.upenn.edu

Date of Publication: December 5, 2012